

**THE KAVERY ENGINEERING COLLEGE***(An Autonomous Institution)***MCA DEGREE END SEMESTER THEORY EXAMINATIONS, NOV/DEC 2025***Third Semester**Master of Computer Applications***PMCE2420- Data Mining and Data Warehousing Techniques***Regulations - 2024**Duration : 3 Hrs**Maximum : 100 Marks***PART - A (ANSWER ALL QUESTIONS) (Marks 10\*2=20)**

| <i>Q.No</i> | <i>Questions</i>                                                  | <i>BTL</i> | <i>CO</i> | <i>Marks</i> |
|-------------|-------------------------------------------------------------------|------------|-----------|--------------|
| 1.          | Define data mining. What are its key objectives?                  | L1         | 1         | 2            |
| 2.          | Differentiate between data cleaning and data reduction.           | L2         | 1         | 2            |
| 3.          | What is an association rule? Give an example.                     | L1         | 2         | 2            |
| 4.          | State the difference between classification and prediction.       | L2         | 2         | 2            |
| 5.          | List any two advantages of decision tree classification.          | L1         | 3         | 2            |
| 6.          | Explain the concept of support vector machines in brief.          | L2         | 3         | 2            |
| 7.          | Define clustering. Name any two partitioning methods.             | L1         | 4         | 2            |
| 8.          | Distinguish between k-Means and k-Medoids clustering.             | L2         | 4         | 2            |
| 9.          | What is a data warehouse? Mention any two of its characteristics. | L1         | 5         | 2            |
| 10.         | Differentiate between OLAP and OLTP systems.                      | L2         | 5         | 2            |

**PART - B (ANSWER ALL QUESTIONS) (Marks 5\*16 = 80)**

| <i>Q.No</i> | <i>Questions</i>                                                                                     | <i>BLT</i> | <i>CO</i> | <i>Marks</i> |
|-------------|------------------------------------------------------------------------------------------------------|------------|-----------|--------------|
| 11.         | a i Describe the different kinds of data handled in data mining with suitable examples.              | L1         | 1         | 8            |
|             | ii Discuss the major challenges and issues involved in data preprocessing.                           | L2         | 1         | 8            |
|             | Or                                                                                                   |            |           |              |
|             | b i Explain the various steps involved in data cleaning and integration.                             | L1         | 1         | 8            |
|             | ii Summarize the importance of data transformation and reduction in improving mining efficiency.     | L2         | 1         | 8            |
| 12.         | a i Illustrate the process of mining frequent item sets using the Apriori algorithm.                 | L2         | 2         | 8            |
|             | ii Apply association rule generation on a small transactional database and explain its significance. | L3         | 2         | 8            |
|             | Or                                                                                                   |            |           |              |
|             | b i Explain the candidate generation and pruning steps in association rule mining.                   | L2         | 2         | 8            |
|             | ii Demonstrate how classification and prediction models are prepared for new data.                   | L3         | 2         | 8            |

|     |    |    |                                                                                                                               |    |   |   |
|-----|----|----|-------------------------------------------------------------------------------------------------------------------------------|----|---|---|
| 13. | a  | i  | Apply the concept of Bayesian classification to a sample dataset and discuss its accuracy.                                    | L3 | 3 | 8 |
|     |    | ii | Examine the role of support vector machines in classification problems.                                                       | L4 | 3 | 8 |
| Or  |    |    |                                                                                                                               |    |   |   |
|     | b  | i  | Analyze the working of k-Nearest Neighbor (k-NN) classification algorithm.                                                    | L3 | 3 | 8 |
|     |    | ii | Compare linear regression and nonlinear regression models with suitable examples.                                             | L4 | 3 | 8 |
| 14. | a  | i  | Analyze the performance of k-Means clustering when the initial centroids are poorly chosen.                                   | L4 | 4 | 8 |
|     |    | ii | Evaluate the effectiveness of hierarchical clustering methods for complex datasets                                            | L5 | 4 | 8 |
|     | Or |    |                                                                                                                               |    |   |   |
|     | b  | i  | Compare partitioning and model-based clustering methods in terms of scalability and interpretability.                         | L4 | 4 | 8 |
| 15. |    | ii | Assess the impact of fuzzy clustering techniques in handling overlapping clusters.                                            | L5 | 4 | 8 |
|     | a  | i  | Design the schema of a data warehouse for a retail organization and justify the choice of schema type.                        | L5 | 5 | 8 |
|     |    | ii | Evaluate the advantages of OLAP operations in business decision-making.                                                       | L6 | 5 | 8 |
|     | Or |    |                                                                                                                               |    |   |   |
|     | b  | i  | Critically analyze the steps involved in ETL (Extraction, Transformation, and Loading) process for building a data warehouse. | L5 | 5 | 8 |
|     |    | ii | Propose a suitable architecture for integrating multiple heterogeneous data sources.                                          | L6 | 5 | 8 |